HNTES Projects 1 and 2

Zhenzhen Yan, Chris Tracy, Malathi Veeraraghavan
University of Virginia and ESnet
Jan 12-13, 2012

Please send feedback/comments to:
mv5g@virginia.edu, ctracy@es.net

Acknowledgment: Thanks to the US DOE ASCR program office for UVA grants DE-SC002350 and DE-SC0007341 and ESnet grant DE-AC02-05CH11231

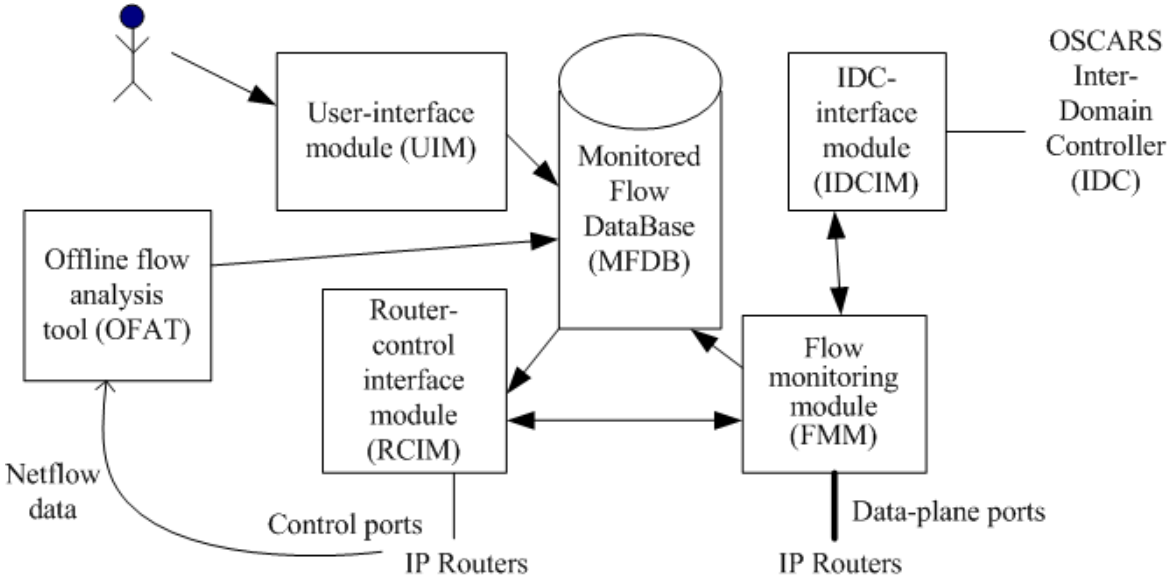
Outline

- Brief history of design work under HNTES project 1 since last DOE PI Meeting in Oct. 2010
- HNTES project 2 work items
  - Completed work: ESnet (1Q Yr 1)
    - ESnet start date: Oct. 1, 2011
  - Planned work: UVA and ESnet
    - UVA start date: Jan. 15, 2012

Brief history since Oct. 2010

- Reported in Oct. 2010 meeting
  - Developed flow analysis algorithms for identifying long duration flows
  - Analyzed Internet2 data (as ESnet data was unavailable to UVA)
  - Demonstrated HNTES 1.0 software
    - Flow Monitoring Module (packets mirrored to this module)
    - IDCIM (IDC interface module)
    - Monitored Flow Data Base (MFDB) – MySQL
  - Focus: dynamic circuit setup

HNTES 1.0 architecture



1. Offline flow analysis and populate MFDB
2. RCIM reads MFDB and programs routers to port mirror packets from MFDB flows
3. Router mirrors packets to FMM
4. FMM asks IDICM to initiate circuit setup as soon as it receives packets from the router corresponding to one of the MFDB flows
5. IDCIM communicates with IDC, which sets up circuit and PBR for flow redirection to newly established circuit

Large size not long duration

- Nov. 2010 - May 2011
  - Changed focus to identifying large sized flows, not long duration flows
  - Why? Because I2 NetFlow analysis showed long duration flows had relatively low rates (33 Mbps)
  - Such flows are not likely to have a significant negative impact on general-purpose flows

Heavy-hitter flows

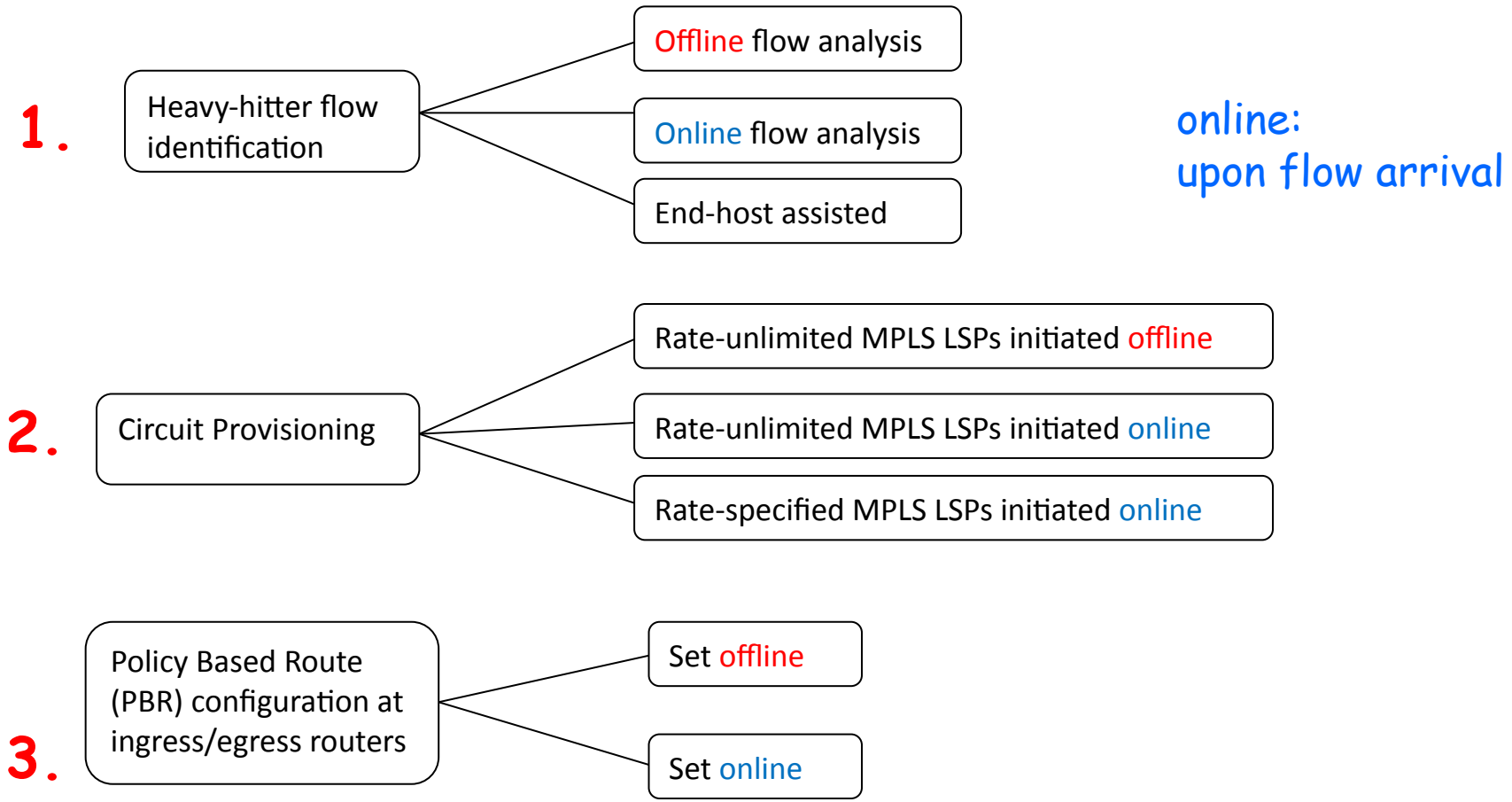
- Dimensions
  - size (bytes): elephant and mice
  - rate: cheetah and snail
  - duration: tortoise and dragonfly
  - burstiness: porcupine (alpha) and stingray (beta)

Kun-chan Lan and John Heidemann, A measurement study of correlations of Internet flow characteristics. *ACM Comput. Netw.* 50, 1 (January 2006), 46-62.

Offline not online circuit prov.

- Nov. 2010 - May 2011
  - But large sized flows were found to be mostly of short duration
    - As will be reported in the next talk
    - Both NetFlow and GridFTP logs
  - Hence changed to offline circuit provisioning and PBR configuration

Three HNTES tasks



HNTES 1.0 vs. HNTES 2.0

	HNTES 1.0 (tested on ANI testbed)	HNTES 2.0
Dimension of heavy-hitter flow	Duration	Size/min
Circuit granularity	Circuit for each flow	Circuit carries multiple flows
Heavy hitter flow identification	Online	Offline
Circuit provisioning	Online	Offline
Flow redirection (PBRconfiguration)	Online	Offline

Focus: DYNAMIC (or online) circuit setup ⇒ IDC circuit setup delay is about 1 minute ⇒ Can use circuits only for long-DURATION flows

HNTES 1.0 logic

UVA-ESnet collaboration started in Mar. 2011

- Feb. 2011: MV attended ESCC meeting and DOE Terabits workshop
  - Steve Cotter's talk on need for science flow redirection
  - Steve said Chris was working on identifying science flows
- March 2011: started collaboration with Chris
- First step:
  - Determine if NetFlow data inspite of 1-1000 sampling is sufficient for offline flow identification of elephant flows
  - Findings presented in first talk
  - Conclusion: NetFlow data is sufficient
  - Hence Flow Monitoring Module (FMM) was dropped

HNTES 2.0: Large sized flow identification algorithm

- Ideally, flow size = total number of bytes sent per flow (5-tuple identifier)
- But since ports are ephemeral, cannot use for offline flow identification
- Redefined "flow": src/dst IP addresses only
- Algorithm
  - Add sizes for a flow from all flow records in say one day
  - Flows with total size > threshold (e.g. 1GB) are monitored
- Combine persistency with size to decide which flow identifiers to include in PBR table

Three events of interest

- May 2011: submitted HNTES Project 2 proposal to DOE
  - At this point, focus: size and offline
- June 2011:
  - Presented a talk on HNTES to ESnet
  - Attendees: Evangelos, Brian, Chin, Eli, Inder, Joe B., Joe M., and Jon
- Aug. 2011:
  - Zhenzhen Yan defended PhD proposal
  - Literature search

What is an "elephant" flow?

- ESnet talk attendees noted:
  - Why 1 day for size aggregation and not 1 hour?
  - Why /32 src/dst IP addr. and not /24?
- Literature:
  - Lan and Heidemann: elephants: total number of bytes in a flow exceeds mean + 3 SD; (total duration is required)
  - Willinger 2005: elephants are the top-ranked flows that send the most number of bytes within 1-minute intervals (once an elephant, always an elephant)
  - Baraniuk 2011: defined alpha flows as flows that exceed a (large) threshold of bytes transmitted in each T-sec bin. Threshold is set as mean + few SD

Elephant vs. alpha

- Heidemann's data analysis:
  - 68% of porcupine flows are elephants (i.e., bursty flows are also large sized)
  - only 19% of elephants are porcupines
- HNTES:
  - adopted Baraniuk's alpha flow definition with the exception that the threshold is set independent of traffic (e.g., 1GB in 1 min)
- Why?
  - Baraniuk and Heidemann: alpha flows are caused by transfers of large files over fast links
  - Baraniuk: traffic bursts typically arise from just a few high-volume connections that dominate all others – such dominating connections are called *alpha traffic.*

Change dimension of heavy-hitter flows

- Change from size to burstiness
  - size (bytes): elephant and mice
  - rate: cheetah and snail
  - duration: tortoise and dragonfly
  - burstiness: porcupine (alpha) and stingray (beta)
- Relation of alpha flows
  - An alpha flow is defined as one in which the bytes generated ≥ H (threshold) in a specified small time interval (e.g., 1 minute: NetFlow active timeout) anytime during its lifetime
    - if H = 1 GB ⇒ throughput exceeds 133 Mbps for any 1 min in lifetime

S. Sarvotham, R. Riedi, and R. Baraniuk, "Connection-level analysis and modeling of nework traffic," in *ACM SIGCOMM Internet Measurement Workshop 2001*, November 2001, pp. 99–104.

HNTES 2.0 June 2011-now

- alpha flow identification
  - as will be presented in the next talk
  - NetFlow data analysis
- GridFTP analysis

ESnet/UVA joint work

- Process:
  - UVA provides NetFlow analysis code + anonymization code to ESnet
  - ESnet executes this code on ESnet NetFlow data, and sends anonymized results to UVA
  - UVA conducts further analysis, generates graphs for papers
- Analyses:
  - NetFlow experiments/analysis: Mar-Apr. 2011
  - GridFTP/NetFlow correlation analysis: Apr.-May 2011
  - Size based NERSC PE router NetFlow analysis: June 2011
  - alpha flow ESnet PE router NetFlow analysis: July 2011-Jan. 2012

Papers/presentations

- Paper at OFC, March 2011
- Heterogeneous net. w/s, March 2011
- I2 Spring member meeting, Apr. 2011
- IEEE Comm. Mag. spl. issue: May 2011
- Talk to ESnet: June 2011
- ICC 2011 paper submission: Sep. 2011
- Traffic engr paper for IEEE HPSR, Jan. 2012 (under prep.)

Outline

- Brief history of design work under HNTES project 1 since last DOE PI Meeting in Oct. 2010
- HNTES project 2 work items
  - Completed work: ESnet (1Q Yr 1)
    - ESnet start date: Oct. 1, 2011
  - Planned work: UVA and ESnet
    - UVA start date: Jan. 15, 2012

HNTES project 2
ESnet 1Q work

- Completed work items
  - Executed alpha flow identification algorithms on 7 months NetFlow data from ESnet site PE router
  - Executed general-purpose flow identification algorithms on same set
  - Collaborated with UVA on two papers

HNTES project 2 Planned work

- HNTES 3.0
  - online flow identification
    - Faster NetFlow data retrieval
    - FMM with 0-length packet mirroring
  - HNTES 4.0
    - end-host assisted flow indentification
    - need PerfSONAR to find route and send independent messages to HNTES in intermediate domains
  - HNTES design for an integrated network
    - unlike ESnet4 with separate IP-routed and SDN networks
    - rate-unlimited MPLS LSPs and third queue concepts
  - Other types of heavy-hitter flows
  - Experiment with OpenFlow on ANI 100G prototype/ LIMAN

Breakdown of work

- UVA
  - Algorithm design and coding
  - Data analysis
  - HNTES software prototyping
- ESnet
  - Review/improve designs
  - Execute scripts on ESnet data
- Both
  - ANI 100G prototype & LIMAN testing

Summary

- HNTES 2.0 offline design appears feasible
- HNTES 3.0 online design challenging
- HNTES 4.0 end-host assisted (builds on Lambdastation/Terapaths)
- Need to design solution for deploying HNTES in an integrated network